Домашний письменный экзамен по курсу математической статистики, 10-15 декабря 2011 г.

Задача 1. Даны две независимые выборки $X_1, \dots, X_n$ и $Y_1, \dots, Y_m$ из бернуллевских распределений $B(p)$ и $B(p')$ соответственно. Предложите тест для проверки гипотезы $H_0 = \{p = p'\}$ против $H_1 = \{p \neq p'\}$ с заданным уровнем значимости. Каким будет результат его применения, если $n = 125$, $m = 50$, $X_1 + \dots + X_{125} = 75$, $X_1 + \dots + X_{50} = 30$, при уровне значимости $\alpha = 5\%$?

Задача 2. а) Пусть $\xi : \Omega \to \mathbb{R}$ — случайная величина, такая, что $\forall x \quad P(\xi(\omega) = x) = 0$, а $F_\xi(x) = P(\xi(\omega) < x)$ — её функция распределения. Как распределена случайная величина $F(\xi)$?
б*) Рассмотрим для независимой выборки $X_1, \dots, X_n$ размера n из распределения в пункте а) соответствующую эмпирическую функцию распределения $\hat{F}(x) := \frac{1}{n}\#\{j \mid X_j \leq x\}$ и статистику Колмогорова

$$\eta = \sup_x |F_\xi(x) - \hat{F}(x)|.$$

Покажите, что распределение статистики η не зависит от исходного распределения F_ξ.
в*) Придумайте применение результату пункта б).

Задача 3. В городе N провели социологический опрос. В выборке из 3100 случайно выбранных человек оказалось 1350 курильщиков. Дайте интервальную оценку для доли курильщиков в городе N с уровнем значимости $\alpha = 2.5\%$.

Задача 4. Из неизвестного распределения, являющегося либо равномерным на $[0, 1]$ (гипотеза H_0), либо имеющим на $[0, 1]$ плотность $\rho(x) = 2x$ (гипотеза H_1), взята независимая выборка x, y размера 2. Постройте оптимальный в смысле теоремы Неймана-Пирсона критерий заданного уровня значимости α. Что он даёт при $x = 0.25$, $y = 0.31$, $\alpha = 5\%$?

Задача 5. Дана выборка размера n из равномерного распределения $R([-\theta, \theta])$. Оцените значение параметра θ методом моментов и методом наибольшего правдоподобия, и найдите матожидание и дисперсию построенных оценок.

Задача 6. Неизвестный параметр θ имеет равномерное распределение на отрезке $[0, 1]$, а соответствующее ему распределение имеет плотность

$$\rho_\theta(x) = \begin{cases} \frac{1}{2}(1 + \theta x), & x \in [-1, 1] \\ 0 & \text{иначе.} \end{cases}$$

По выборке из а) одного б) двух независимых экспериментов предложите оптимальную в смысле байесовского подхода с квадратичной функцией риска $R(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$ оценку параметра θ.

Задача 7. Даны две независимые выборки $X_1, \dots, X_n$ и $Y_1, \dots, Y_m$ из распределений $\mathcal{N}(a, \sigma^2)$ и $\mathcal{N}(a', (\sigma')^2)$ соответственно. Предложите тест для проверки гипотезы $H_0 = \{a = a'\}$ против $H_1 = \{a \neq a'\}$ с заданным уровнем значимости (при этом можно ссылаться на квантили не выписанного явно, но чётко заданного, распределения). Что можно сказать о случае выборок

$$1.076740, 0.560665, 1.607037, 1.977505, 1.564703, 2.116484, 0.560824,$$

$$0.114073, -0.031772, -1.834089, 1.364763, 0.537780, 0.581747, 1.062363,$$

$$1.387948, 1.795011, 0.069909, 1.778455, 0.172072, 0.838818, 1.611839$$

и

$$1.604971, 0.006220, 0.106422, 0.461170, 2.146923, -0.354759, 0.064304,$$

$$-0.798859, 2.596132, 0.830454, 1.229313, -0.350428, -0.345275, 0.231978,$$

$$-0.307687, -2.082400, -1.149044, 0.368416, -0.332208, 0.602481, -0.014309,$$

при уровне значимости $\alpha = 5\%$?